Nested and Random Effects Models[1]

STA441 Winter 2016

[1]See last slide for copyright information.

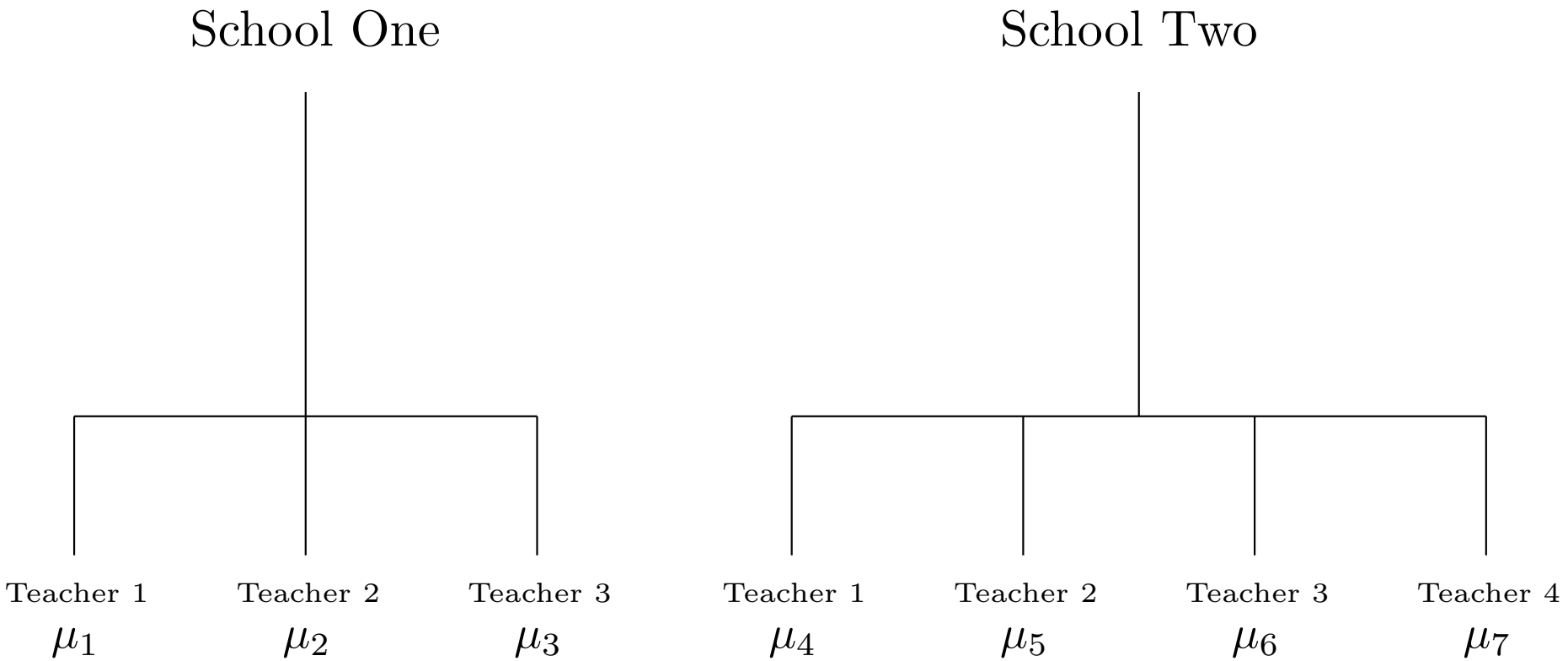
Overview

Nested Designs
Example

A chain of commercial business colleges is teaching a software certification course. After 6 weeks of instruction, students take a certification exam and receive a score ranging from zero to 100.

- The owners want to see whether performance is related to which school students attend, or which instructor they have – or both.
- They compare two schools; one of the schools has three instructors teaching the course, and the other school has 4 instructors teaching the course.
- A teacher only works in one school.
- There are two categorical explanatory variables, school and teacher.
- But it's not a factorial design, because "Teacher 1" does not mean the same thing in School 1 and School 2.
- It's a different person.

Teacher is nested within school



Schools $H_0 : \frac{1}{3}(\mu_1 + \mu_2 + \mu_3) = \frac{1}{4}(\mu_4 + \mu_5 + \mu_6 + \mu_7)$

Teachers within Schools $H_0 : \mu_1 = \mu_2 = \mu_3$ and $\mu_4 = \mu_5 = \mu_6 = \mu_7$

Tests of nested effects are tests of contrasts

$H_0 : \frac{1}{3}(\mu_1 + \mu_2 + \mu_3) = \frac{1}{4}(\mu_4 + \mu_5 + \mu_6 + \mu_7)$
$H_0 : \mu_1 = \mu_2 = \mu_3$ and $\mu_4 = \mu_5 = \mu_6 = \mu_7$

You can specify the contrasts yourself, or you can take advantage of `proc glm`'s syntax for nested models.

```
proc glm;
     class school teacher;
     model score = school teacher(school);
```

The notation `teacher(school)` should be read "teacher within school."

Easy to extend the ideas

- Can have more than one level of nesting. You could have climate zones, lakes within climate zones, fishing boats within lakes, ...
- There is no problem with combining nested and factorial structures. You just have to keep track of what's nested within what.
- Factors that are not nested are sometimes called "crossed."
- The combination of nesting and *random effects* is very powerful.

Random Effects

As opposed to *fixed effects*

A random factor is one in which the *values of the factor are a random sample* from a populations of values.

- Randomly select 10 schools, test students at each school. School is a random factor with 10 values.
- Randomly select 15 naturopathic medicines for arthritis (there are quite a few), and then randomly assign arthritis patients to try them. Drug is a random factor.
- Randomly select 15 lakes. In each lake, measure how clear the water is at 20 randomly chosen points. Lake is a random factor.
- Randomly select 20 fast food outlets, survey customers in each about quality of the fries. Outlet is a random factor with 20 values. Amount of salt would be a fixed factor, which could be crossed with outlet.

One random factor

A nice simple example

- Randomly select 5 farms.
- Randomly select 10 cows from each farm, milk them, and record the amount of milk from each one.
- The one random factor is Farm.
- Total $n = 50$.
- The idea is that "Farm" is a kind of random shock that pushes all the amounts of milk in a particular farm up or down by the same amount.
- You could also think of cow (the cases are cows) as a random factor nested within farm.

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

Analysis of variance

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

$$\begin{aligned} Var(Y_{ij}) &= Var(\mu + \tau_i + \epsilon_{ij}) \\ &= Var(\tau_i) + Var(\epsilon_{ij}) \\ &= \sigma_\tau^2 + \sigma^2 \end{aligned}$$

- Split the variance up into two parts: The part that comes from farms, and the part that comes from cows (within farms).
- *Analysis* of variance.
- Test $H_0 : \sigma_\tau^2 = 0$
- Estimate $\frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma^2}$

Nesting and random effects

- Nested models are often viewed as random effects models, but there is no necessary connection between the two concepts.
- It depends on how the study was conducted. Were the two schools randomly selected from some population of schools, or did someone just pick those two (maybe because there are just two schools)?
- Random effects, like fixed effects, can either be nested or not; it depends on the logic of the design.

Sub-sampling

- Sub-sampling is a useful case of nested and purely random effects.
- For example,
  - Select a random sample of towns.
  - From each town, select a random sample of households.
  - From each household, select a random sample of people to test, or measure, or question.
- Start with the grand mean μ.
- Town, household and person are random shocks that push the measurement up or down from the grand mean.
- In the model, shocks are independent normal with mean zero.
- Analysis of variance.

Sub-sampling with SAS

Waste water treatment

- We are studying the porosity of "flocks," nasty little pieces of something floating in the tanks.
- We randomly select a sample of flocks, and then cut each one up into very thin slices.
- We then randomly select a sample of slices (called "sections") from each flock, look at it under a microscope, and assign a number representing how porous it is (how much empty space there is in a designated region of the section).
- The explanatory variables are flock and section.
- The research question is whether section is explaining a significant amount of the variance in porosity – because if not, we can use just one section per flock, and save considerable time and expense.

SAS `proc nested`

SAS `proc nested` is built specifically for pure random effects models with each explanatory variable nested within all the preceding ones.

```
proc sort; by flock section; /* Data must be sorted */
proc nested;
     class flock section;
     var por;
```

You could use `proc glm`, but the `proc nested` syntax is easier and the output is nicer for this special case.

Mixed models
The classical approach

- There can be both fixed and random factors in the same experiment. This makes it a *mixed* model.
- Factors can be nested or crossed, in various patterns.
- Random factors can be nested within fixed.
- Fixed effects cannot be nested within random.
- The interaction of any random factor with another factor (whether fixed or random) is random.
- F-tests are often possible, but they don't always use Mean Squared Error in the denominator of the F statistic.
- Often, it's the Mean Square for some interaction term.
- The choice of what error term to use is relatively mechanical for balanced models — based on expected mean squares.
- Mechanical means SAS can do it for you.

One more example
And some sample questions

Independent random samples of 10 Canadian and 10 U.S. large companies were selected. In each company, 25 female and 25 male managers were randomly selected, and their formal education in years was recorded.

1. Is this an observational study, or experimental? Observational.
2. What are the factors? Nation, Company and Sex.
3. Designate the factors as fixed or random. Nation and Sex are fixed. Company is random.
4. Describe the nesting, if any. Company is nested within Nation.

Copyright Information